

Be
#10
PCT09

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001
TIME: 11:43:34

Input Set : A:\ES.txt
Output Set: N:\CRF3\08302001\I807867.raw

ENTERED

3 <110> APPLICANT: CAROL, Pierre
4 KUNTZ, Marcel
5 MACHE, Regis
7 <120> TITLE OF INVENTION: cDNA SEQUENCE TRANSCRIBING AN mRNA ENCODING THE TERMINAL
OXIDASE
8 ASSOCIATED WITH CAROTENOID BIOSYNTHESIS, AND USES THEREOF
10 <130> FILE REFERENCE: 109326
12 <140> CURRENT APPLICATION NUMBER: US 09/807,867
13 <141> CURRENT FILING DATE: 2001-06-15
15 <150> PRIOR APPLICATION NUMBER: PCT/IB99/01719
16 <151> PRIOR FILING DATE: 1999-10-20
18 <150> PRIOR APPLICATION NUMBER: FR 9813283
19 <151> PRIOR FILING DATE: 1998-10-20
21 <160> NUMBER OF SEQ ID NOS: 10
23 <170> SOFTWARE: PatentIn version 3.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 1396
27 <212> TYPE: DNA
28 <213> ORGANISM: Arabidopsis thaliana
30 <400> SEQUENCE: 1
31 cgcgtcacat tgggattcgt cattcttctt ctaaaaccgc caaaatttct ccattttctac 60
33 caaaaatata caacttttac ttttctttcc tgtgaaatta tctgctcaaa tctttgggtc 120
35 ctgacggaga tggcggcgat ttcaggcatc tcctctggta cgttgacgat ttcacggcct 180
37 ttggttactc ttcgacgctc tagagccgcc gtttcgtaca gctcctctca ccgattgctt 240
39 catcatcttc ctctctcttc tgcgtcgtctg ctattaagga acaatcatcg agtccaagca 300
41 acgattttgc aagacgatga agagaaagtg gtggtggagg aatcggttaa agccgagact 360
43 tctactggta cagaaccact tgaggagcca aatatgagtt cttcttcaac tagtgctttt 420
45 gagacatgga tcatcaagct tgagcaagga gtgaatgttt tccttacaga ctcggttatt 480
47 aagatacttg acactttgta tcgtgaccga acatatgcaa ggttctttgt tcttgagaca 540
49 attgctagag tgccttattt tgcgtttatg tctgtgctac atatgtatga gacctttggt 600
51 tgggtggagg gagcagatta tttgaaagta cactttgctg agagctggaa tgaaatgcat 660
53 cacttgctca taatggaaga attgggtgga aattcttggt ggtttgatcg ttttctggct 720
55 cagcacatag caaccttcta ctacttcatg acagtgttct tgtatatctt aagccctaga 780
57 atggcatatc acttttccga atgtgtggag agtcatgcat atgagactta tgataaattt 840
59 ctcaaggcca gtggagagga gttgaagaat atgcctgcac cggatatcgc agtaaaatac 900
61 tatacgggag gtgacttgta cttatttgat gagttccaaa catcaagaac tcccaatact 960
63 cgaagaccag taatagaaaa tctatacgat gtgttttgta acataagaga tgatgaagca 1020
65 gaacactgca agacaatgag agcttgtcag actctaggca gtctgcgttc tccacactcc 1080
67 attttagatg atgatgatac tgaagaagaa tcagggtgtg ttgttctcga ggaggctcat 1140
69 tgcgaaggta ttgtagactg cctcaagaaa tccattacaa gttaataaat tagaaagtaa 1200
71 actaaaaaag attatttgta tcagctcatg aacaatagat ataatcccat atacttgga 1260
73 ataaaggaat aatgtgaaat tcccatcggt gtgctagtgt gtgagagaat caaataccct 1320
75 aatgatgtaa atgtactttg atgagcttaa gtcgtttag accattttat caaaaaaaaa 1380
77 aaaaaaaaaa aaaaaa
80 <210> SEQ ID NO: 2
81 <211> LENGTH: 351
82 <212> TYPE: PRT
83 <213> ORGANISM: Arabidopsis thaliana

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001

TIME: 11:43:34

Input Set : A:\ES.txt

Output Set: N:\CRF3\08302001\I807867.raw

85 <400> SEQUENCE: 2

```

87 Met Ala Ala Ile Ser Gly Ile Ser Ser Gly Thr Leu Thr Ile Ser Arg
88 1          5          10          15
90 Pro Leu Val Thr Leu Arg Arg Ser Arg Ala Ala Val Ser Tyr Ser Ser
91          20          25          30
93 Ser His Arg Leu Leu His His Leu Pro Leu Ser Ser Arg Arg Leu Leu
94          35          40          45
96 Leu Arg Asn Asn His Arg Val Gln Ala Thr Ile Leu Gln Asp Asp Glu
97          50          55          60
99 Glu Lys Val Val Val Glu Glu Ser Phe Lys Ala Glu Thr Ser Thr Gly
100 65          70          75          80
102 Thr Glu Pro Leu Glu Glu Pro Asn Met Ser Ser Ser Ser Thr Ser Ala
103          85          90          95
105 Phe Glu Thr Trp Ile Ile Lys Leu Glu Gln Gly Val Asn Val Phe Leu
106          100          105          110
108 Thr Asp Ser Val Ile Lys Ile Leu Asp Thr Leu Tyr Arg Asp Arg Thr
109          115          120          125
111 Tyr Ala Arg Phe Phe Val Leu Glu Thr Ile Ala Arg Val Pro Tyr Phe
112          130          135          140
114 Ala Phe Met Ser Val Leu His Met Tyr Glu Thr Phe Gly Trp Trp Arg
115 145          150          155          160
117 Arg Ala Asp Tyr Leu Lys Val His Phe Ala Glu Ser Trp Asn Glu Met
118          165          170          175
120 His His Leu Leu Ile Met Glu Glu Leu Gly Gly Asn Ser Trp Trp Phe
121          180          185          190
123 Asp Arg Phe Leu Ala Gln His Ile Ala Thr Phe Tyr Tyr Phe Met Thr
124          195          200          205
126 Val Phe Leu Tyr Ile Leu Ser Pro Arg Met Ala Tyr His Phe Ser Glu
127          210          215          220
129 Cys Val Glu Ser His Ala Tyr Glu Thr Tyr Asp Lys Phe Leu Lys Ala
130 225          230          235          240
132 Ser Gly Glu Glu Leu Lys Asn Met Pro Ala Pro Asp Ile Ala Val Lys
133          245          250          255
135 Tyr Tyr Thr Gly Gly Asp Leu Tyr Leu Phe Asp Glu Phe Gln Thr Ser
136          260          265          270
138 Arg Thr Pro Asn Thr Arg Arg Pro Val Ile Glu Asn Leu Tyr Asp Val
139          275          280          285
141 Phe Val Asn Ile Arg Asp Asp Glu Ala Glu His Cys Lys Thr Met Arg
142          290          295          300
144 Ala Cys Gln Thr Leu Gly Ser Leu Arg Ser Pro His Ser Ile Leu Asp
145 305          310          315          320
147 Asp Asp Asp Thr Glu Glu Glu Ser Gly Cys Val Val Pro Glu Glu Ala
148          325          330          335
150 His Cys Glu Gly Ile Val Asp Cys Leu Lys Lys Ser Ile Thr Ser
151          340          345          350

```

153 <210> SEQ ID NO: 3

154 <211> LENGTH: 1387

155 <212> TYPE: DNA

156 <213> ORGANISM: capsicum

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001
TIME: 11:43:34

Input Set : A:\ES.txt
Output Set: N:\CRF3\08302001\I807867.raw

```

158 <400> SEQUENCE: 3
159 ccacgcgtcc gataaaaaaa tcaagaatgg cgattttccat atctgctatg agttttcgaa 60
161 cttcagtttc ttcttcatat tcagcatttt tgtgcaattc caagaaccca tttgtttga 120
163 attctctatt ttacttagg aattctcata gaacttttca gccttcgtta tcaaggaaat 180
165 caagtagagt tcgagcaacg ttgttaaaag agaatgaaga agaagtgggt gtggagaaat 240
167 cttttgcacc taagagtttt cctggtaatg tgggaggggg aaataatggg gagccacccg 300
169 ataattcatc ctogaacggt ctggagaaat gggttataaa gattgagcag tctgtaaata 360
171 tctttctcac ggattcagtg ataaagattc ttgacacttt gtatcacgac cgacactatg 420
173 cgaggttttt cgttctggaa acaattgcaa gaggttcctta ttttgcattt atatctgttc 480
175 ttacttgta cgagagcttt gggttggtgga gacgagcaga ttatctgaag gtgcattttg 540
177 ccgagagctg gaatgagatg caccattttac tcattatgga ggaattaggt ggaaatgctt 600
179 ggtggtttga ccgattcctt gcgcaacata ttgctgtatt ctattatttc atgacagtct 660
181 cgatgtatgc tttgagcccg agaatggcat atcattttct tgaatgtgtg gagcaccatg 720
183 catacgagac ttacgataaa ttcatcaagg atcaagaagc ggaattgaag aaattgcccg 780
185 ctccaaagat tgcagtgagc tactacaccg gaggtgactt gtatttgttc gatgagtttc 840
187 aaacatcacg agagcctaact actogaaggc caaaaataga taatctgtac gacgtattca 900
189 tgaacatcag agatgacgaa gcagagcatt gtaagacaat gaaagcgtgt caaaccocatg 960
191 ggagcctccg ctcccctcac acaaatccat gcgatgagtc tgaagacgat ccaggttggt 1020
193 cagtgcctca ggccgattgt gtaggtatcg tggattgtat aacgaaatct gtcgctgatc 1080
195 ctaacgctcg cagaaggtag ggaaaggaaa aacgcagaac gaaactatac atgtatatac 1140
197 cagtacagcc aaatatacaa gaaatataca tacatattgt atcttttact ctctgaggaa 1200
199 gagcttgta aattgcccaa aaaatgggta ggcacttggt tttgttttca cctttcaata 1260
201 atttgtaact aactatgaac aaatttgctc cggcacacta caactccata ggggtcctgt 1320
203 tacgcttctg aactaaattt taacatatit ttgtcaacct tctcagcaaa aaaaaaaaaa 1380
205 aaaaaaa
208 <210> SEQ ID NO: 4
209 <211> LENGTH: 357
210 <212> TYPE: PRT
211 <213> ORGANISM: capsicum
213 <400> SEQUENCE: 4
215 Met Ala Ile Ser Ile Ser Ala Met Ser Phe Arg Thr Ser Val Ser Ser
216 1 5 10 15
218 Ser Tyr Ser Ala Phe Leu Cys Asn Ser Lys Asn Pro Phe Cys Leu Asn
219 20 25 30
221 Ser Leu Phe Ser Leu Arg Asn Ser His Arg Thr Phe Gln Pro Ser Leu
222 35 40 45
224 Ser Arg Lys Ser Ser Arg Val Arg Ala Thr Leu Leu Lys Glu Asn Glu
225 50 55 60
227 Glu Glu Val Val Val Glu Lys Ser Phe Ala Pro Lys Ser Phe Pro Gly
228 65 70 75 80
230 Asn Val Gly Gly Gly Asn Asn Gly Glu Pro Pro Asp Asn Ser Ser Ser
231 85 90 95
233 Asn Gly Leu Glu Lys Trp Val Ile Lys Ile Glu Gln Ser Val Asn Ile
234 100 105 110
236 Phe Leu Thr Asp Ser Val Ile Lys Ile Leu Asp Thr Leu Tyr His Asp
237 115 120 125
239 Arg His Tyr Ala Arg Phe Phe Val Leu Glu Thr Ile Ala Arg Val Pro
240 130 135 140
242 Tyr Phe Ala Phe Ile Ser Val Leu His Leu Tyr Glu Ser Phe Gly Trp

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001

TIME: 11:43:34

Input Set : A:\ES.txt

Output Set: N:\CRF3\08302001\I807867.raw

243	145				150				155				160			
245	Trp	Arg	Arg	Ala	Asp	Tyr	Leu	Lys	Val	His	Phe	Ala	Glu	Ser	Trp	Asn
246					165				170						175	
248	Glu	Met	His	His	Leu	Leu	Ile	Met	Glu	Glu	Leu	Gly	Gly	Asn	Ala	Trp
249				180					185					190		
251	Trp	Phe	Asp	Arg	Phe	Leu	Ala	Gln	His	Ile	Ala	Val	Phe	Tyr	Tyr	Phe
252			195					200					205			
254	Met	Thr	Val	Ser	Met	Tyr	Ala	Leu	Ser	Pro	Arg	Met	Ala	Tyr	His	Phe
255		210					215					220				
257	Ser	Glu	Cys	Val	Glu	His	His	Ala	Tyr	Glu	Thr	Tyr	Asp	Lys	Phe	Ile
258	225					230				235					240	
260	Lys	Asp	Gln	Glu	Ala	Glu	Leu	Lys	Lys	Leu	Pro	Ala	Pro	Lys	Ile	Ala
261					245					250					255	
263	Val	Ser	Tyr	Tyr	Thr	Gly	Gly	Asp	Leu	Tyr	Leu	Phe	Asp	Glu	Phe	Gln
264					260				265					270		
266	Thr	Ser	Arg	Glu	Pro	Asn	Thr	Arg	Arg	Pro	Lys	Ile	Asp	Asn	Leu	Tyr
267			275					280						285		
269	Asp	Val	Phe	Met	Asn	Ile	Arg	Asp	Asp	Glu	Ala	Glu	His	Cys	Lys	Thr
270		290				295						300				
272	Met	Lys	Ala	Cys	Gln	Thr	His	Gly	Ser	Leu	Arg	Ser	Pro	His	Thr	Asn
273	305					310				315					320	
275	Pro	Cys	Asp	Glu	Ser	Glu	Asp	Asp	Pro	Gly	Cys	Ser	Val	Pro	Gln	Ala
276				325						330					335	
278	Asp	Cys	Val	Gly	Ile	Val	Asp	Cys	Ile	Thr	Lys	Ser	Val	Ala	Asp	Pro
279			340						345					350		
281	Asn	Val	Gly	Arg	Arg											
282			355													
284	<210>	SEQ ID NO:	5													
285	<211>	LENGTH:	1284													
286	<212>	TYPE:	DNA													
287	<213>	ORGANISM:	tomato													
289	<400>	SEQUENCE:	5													
290	gaattcggca	cgagcggcac	gagcagaaaa	ctaacaactt	tcccactttg	gaattttctt									60	
292	taccttacct	aagaagggtta	ttaatttgat	tcttggtgga	aggaagaagg	atcaagaatg									120	
294	gcgatttcga	tttctgctat	gagttttgga	acctcagttt	cttcatattc	ttgttttaga									180	
296	gctaggagtt	ttgagaagtc	atcagtttta	tgcaattccc	agaacccatg	tcggtttaat									240	
298	tctgtttttc	cgattcggaa	atctgatggg	gcttcacggt	gttctgtttc	taggaaatca									300	
300	tgtagagttc	gagcaacggt	gttacaagag	aatgaagaag	aagtgggtgt	ggagaaatct									360	
302	tttgacaccta	agagttttcc	tgataacgtg	ggagggggaa	gtaatgggaa	gccaccagat									420	
304	gattcatcct	ctaacggtct	agagaaatgg	gttataaagc	ttgagcagtc	tgtaaataatc									480	
306	ttactcacgg	attcagtgat	aaagattcct	gacactttgt	atcacaaccg	aaactatgcg									540	
308	aggttttttg	ttctggaaac	aattgcaagg	gttccttatt	ttgcatttat	atcggttctt									600	
310	cacatgtatg	agagctttgg	ctggtggaga	agggcagatt	atatgaaaag	gcattttgct									660	
312	gaaagctgga	atgagatgca	ccatttgctc	attatggaag	aattaggggg	aaatgcttgg									720	
314	tggtttgatc	gatttcttgc	acaacatata	gctatattct	attatttcat	gacagtcttg									780	
316	atgtatgctt	tgagcccag	aatggcatat	catcttctctg	aatgtgtgga	gagccatgca									840	
318	tacgagactt	acgataaatt	catcaaggat	caaggagagg	aattgaagaa	tttgcccgct									900	
320	ccaaagattg	cagtggacta	ctacacggga	ggtgacttat	atttatttga	tgagtttcaa									960	
322	acttcacgag	agcctaatac	tcgaagacca	aaaatagata	atctctatga	cgtatttcag									1020	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001
TIME: 11:43:34

Input Set : A:\ES.txt
Output Set: N:\CRF3\08302001\I807867.raw

```

324 aacattagag atgacgaagc agagcattgt aaaacgatga aagcctgtca aactcacggg 1080
326 agccttcggt ctccacacac agatccatgc gatgattctg aagatgatac aggggtgtcc 1140
328 gtacctcaag ctgattgtat aggtatcgtg gattgtataa agaagtcagt caccgatact 1200
330 caagtaacca aaaggttagga aaaggaaaaa cgcggacaaa ctatacttgt atatactagt 1260
332 atagacaaaa aaaaaaaaaa aaaa 1284
335 <210> SEQ ID NO: 6
336 <211> LENGTH: 19
337 <212> TYPE: DNA
C--> 338 <213> ORGANISM: Artificial
340 <220> FEATURE:
341 <223> OTHER INFORMATION: PCR primer ✓
343 <400> SEQUENCE: 6 19
344 gcaacgattt tgcaagacg
347 <210> SEQ ID NO: 7
348 <211> LENGTH: 24
349 <212> TYPE: DNA
C--> 350 <213> ORGANISM: Artificial
352 <220> FEATURE:
353 <223> OTHER INFORMATION: PCR primer ✓
355 <400> SEQUENCE: 7 24
356 ttaacttgta atggatttct tgag
359 <210> SEQ ID NO: 8
360 <211> LENGTH: 171
361 <212> TYPE: PRT
362 <213> ORGANISM: soybean
364 <400> SEQUENCE: 8
366 Tyr Arg Thr Val Lys Leu Leu Arg Ile Pro Thr Asp Leu Phe Phe Lys
367 1 5 10 15
369 Arg Arg Tyr Gly Cys Arg Ala Met Met Leu Glu Thr Val Ala Ala Val
370 20 25 30
372 Pro Gly Met Val Gly Gly Met Leu His Leu Arg Ser Leu Arg Lys
373 35 40 45
375 Phe Gln Gln Ser Gly Gly Trp Ile Lys Ala Leu Leu Glu Glu Ala Glu
376 50 55 60
378 Asn Glu Arg Met His Leu Met Thr Met Val Glu Leu Val Lys Pro Lys
379 65 70 75 80
381 Trp Tyr Glu Arg Leu Leu Val Leu Ala Val Gln Gly Val Phe Phe Asn
382 85 90 95
384 Ala Phe Phe Val Leu Tyr Ile Leu Ser Pro Lys Val Ala His Arg Ile
385 100 105 110
387 Val Gly Tyr Leu Glu Glu Glu Ala Ile His Ser Tyr Thr Glu Tyr Leu
388 115 120 125
390 Lys Asp Leu Glu Ser Gly Ala Ile Glu Asn Val Pro Ala Pro Ala Ile
391 130 135 140
393 Ala Ile Asp Tyr Trp Arg Leu Pro Lys Asp Ala Arg Leu Lys Asp Val
394 145 150 155 160
396 Ile Thr Val Ile Arg Ala Asp Glu Ala His His
397 165 170
399 <210> SEQ ID NO: 9

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/807,867

DATE: 08/30/2001

TIME: 11:43:35

Input Set : A:\ES.txt

Output Set: N:\CRF3\08302001\I807867.raw

L:338 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:6
L:350 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:7